

Assignment 4
CS 510 Information Retrieval on the Internet
8 points; due Tuesday, 6 February

Goal: This assignment is to reinforce concepts related to inverted indexes and similarity calculations.

Details: The exercises below involve Sample Documents d1-d5 from slide 19 of Lecture 6, and the formulas given at (1) on page 6 of the Zobel and Moffat paper.

- A. Compute the document weight W_d for documents d1, d2, d4 and d5.
- B. Show the index entries (as on slide 21) for the terms `like`, `mouse` and `them`.
- C. Compute $S_{q,d}$ for documents d1-d5 on each of the following queries. For each $S_{q,d}$, show the factors $\sum w_{d,t} * w_{q,t}$, W_d and W_q , as well as the final result.
- i. `{them}`
 - ii. `{a, mouse}`
 - iii. `{you, like, mouse}`
- D. Redo the $S_{q,d}$ computations for the query `{you, like, mouse}` assuming `you` is a stopword (and none of the other terms is). Show all component factors of the computation, as in Part C. (Note that all of these factors are affected by this change.)

Useful tip: Try typing $\ln(1 + 5/3)$ into the search box on Google.